Preliminary data of a HAMD-17 validated symptom scale derived from the ICD-10 to diagnose depression in outpatients

Melzer, J ; Rostock, M ; Brignoli, R ; Keck, M E ; Saller, R

DOI: https://doi.org/10.1159/000342018

Forschende Komplementärmedizin
Wissenschaft • Praxis • Perspektiven

Original Article · Originalarbeit

Forsch Komplementmed 2012;19:191–196
DOI: 10.1159/000342018

Published online: August 20, 2012

Preliminary Data of a HAMD-17 Validated Symptom Scale Derived from the ICD-10 to Diagnose Depression in Outpatients

Jörg Melzer[a,b] Matthias Rostock[a] Reto Brignoli[d] Martin E. Keck[c,e] Reinhard Saller[a]

[a] Institute of Complementary Medicine,
[b] Clinic for Psychiatry and Psychotherapy, University Hospital Zurich,
[c] Clienia Schloessli, Private Hospital for Psychiatry and Psychotherapy, Oetwil am See,
[d] Clinical Pharmacology, Tradyser Inc., Rüti,
[e] Neuroscience Center, Zurich, Switzerland

Keywords
ICD-10 score · HAMD-17 · Depression · Affective disorder

Summary
Background: In outpatient settings diagnostic classification of depressive symptoms is mostly descriptive based on ICD-10. Depending on clinical experience and consultation time, diagnosis can be verified by validated scales. However, physicians working in primary care are familiar with ICD-10 criteria. Therefore, the aim of the present study was to examine the feasibility of the validation of an ICD-10-derived symptom scale for depression. **Methods:** For this preliminary trial we generated a symptom scale derived 1:1 from the diagnostic criteria for depression given in the ICD-10 with 10 items. The Hamilton Rating Scale for Depression (HAMD-17) was used as reference in a population of 226 outpatients suffering from depressive symptoms. Correlation between scales as well as sensitivity and specificity of the ICD-10 scale were calculated. **Results:** The generated ICD-10 symptom scale for depression could be analyzed in 219 patients and showed a significant and strong correlation with the HAMD-17 ($p < 0.0001$; $\rho = 0.75$). The best trade-offs between specificity and sensitivity of the ICD-10 score were found at 10 points for the lower and 14 points for the upper cut-off. Overall sensitivity and specificity was 76.7 and 88.6%. Almost two thirds (i.e. 65.3%) of the patients were correctly classified by the ICD-10 scale. **Conclusion:** The ICD-10 symptom scale examined in the current population was found to have fair correlation with the HAMD-17 as well as, in face of the limited variance of the patients' condition, acceptable sensitivity and specificity. Therefore, this preliminary study showed that the ICD-10-derived symptom scale seems appropriate to be investigated in a thorough validation trial.

Schlüsselwörter
ICD-10-Score · HAMD-17 · Depression · Affektive Störung

Zusammenfassung
Hintergrund: Im ambulanten Medizinbereich erfolgt die diagnostische Einteilung depressiver Symptome meist deskriptiv nach ICD-10. Die Diagnose kann in Abhängigkeit von klinischer Erfahrung und Konsultationsdauer mit etablierten Messinstrumenten verifiziert werden. Allerdings sind Kliniker mit dem ICD-10 vertraut. Daher war es das Ziel der vorliegenden Studie, die Durchführbarkeit der Validierung einer nach ICD-10-Kriterien generierten Skala zu analysieren. **Methode:** Für eine vorläufige Studie wurde anhand der diagnostischen Kriterien für Depression im ICD-10 1:1 eine Symptomskala generiert. Die Korrelation wurde mittels des HAMD-17 als Referenzskala an 226 ambulanten Patienten analysiert. Zusätzlich wurden Sensitivität und Spezifität als auch die Cut-off-Werte berechnet. **Ergebnisse:** Die generierte ICD-10-Symptomskala für depressive Symptome konnte für 219 Patienten ausgewertet werden und zeigte eine signifikante und starke Korrelation mit dem HAMD-17 ($p < 0{,}0001$; $\rho = 0{,}75$). Die besten «Trade-Offs» zwischen Spezifität und Sensitivität lagen bei 10 Punkten für den unteren und 14 Punkten für den oberen Cut-Off. Insgesamt betrugen Sensitivität und Spezifität 76,7 und 88,6%. Fast zwei Drittel (65,3%) der Patienten wurde mit der ICD-10-Skala korrekt klassifiziert. **Schlussfolgerung:** Die untersuchte ICD-10-Symptomskala zeigte in der vorliegenden Population gute Ergebnisse bezüglich Korrelation mit dem HAMD-17 sowie Sensitivität und Spezifität. Daher zeigt diese vorläufige Studie, dass sich diese ICD-10-Symptomskala für eine umfassende Validierungsstudie eignet.

KARGER
Fax +49 761 4 52 07 14
Information@Karger.de
www.karger.com

1661-4119/12/0194-0191$38.00/0

Accessible online at:
www.karger.com/fok

Dr. Jörg Melzer
Institute of Complementary Medicine
University Hospital Zurich
Raemistrasse 100, 8091 Zurich, Switzerland
Tel. +41 44 255-2460, Fax -4394
joerg.melzer@usz.ch

Introduction

The prevalence of depression in clinics and in general practices is around 10–25% for women, 5–12% for men and 75–80% for patients experiencing recurrent depression [1–3]. Consequently it is most likely that general practitioners (GPs) and specialists face patients with depressive symptoms among their weekly outpatients.

Yet, GPs are said to fail to diagnose 50% of depressed outpatients [4]. Even in research the screening for depression with valid instruments has demonstrated only limited benefit [5]. In daily practice one might doubt if a lack of knowledge or availability of numerous valid instruments [6–11], widely known among specialists, is the problem or rather the lack of time needed for the correct administration of instruments in the face of limited resources. Yet, one would have to admit that even among psychiatrists the discussion which questionnaire should be used remains lively discussed [12, 13].

Nevertheless, patients, physicians, and health insurances are interested in quality of medical care, although from different viewpoints. Next to the effort to improve therapy according to evidence-based medicine (EbM), the evaluation of the efficacy of treatment is important for patients, physicians and increasingly also for health insurances. Notwithstanding, the first impression of the physician during clinical examination is one of the earliest diagnostic criteria to find out the medical state of the patient.

Unfortunately, the increasing economic pressure has entered the medical system as well, and although this rises many ethical questions [14], time for the patient is often short and the administration of time-consuming validated rating scales seems often hardly possible [7, 12]. Yet, for classification of depressive symptoms in daily practice and for communication between physicians the ICD-10 is a worldwide used system [15]. Within the ICD-10 the operational diagnostic criteria for depressive syndromes are a special case as they cover criteria mentioned in existing validated questionnaires. They have been applied in evalutating outpatients, showed good concordance in diagnosis between GPs as well as psychiatrists [16], and correlated well with the Diagnostic and Statistical Manual for Mental Disorders DSM-III-R [17] or the Symptom Checklist SCL-90 [7].

Therefore, it was of particular interest for us to examine the feasibility in a population of outpatients with depressive syndrome, whether and if so to which extent the symptoms listed in the ICD-10 correlate with a validated instrument to corroborate the rationale for thorough validation trial. The data presented herein are preliminary and investigate the correlation with another validated scale widely used in patients with depressive symptoms, the Hamilton Rating Scale for Depression in its 17-item version (HAMD-17). The results of the open trial on efficacy and safety of the intervention with a drug from St. John's wort in a community-based outpatient setting (n = 1,541) in which the ICD-10-derived scale had also been tested have been published elsewhere [18].

Methods

In a subpopulation of an open study [18], 40 investigators (i.e. GPs, neurologists, psychiatrists) used HAMD-17 next to the constructed ICD-10 scale to evaluate the feasibility for validation. Patients had to be ≥ 18 years old, understand German sufficiently, and give consent to participate.

The reference scale, the HAMD-17, is a multiple choice questionnaire with 17 questions for detecting symptoms of major depression such as depressed mood, guilt, suicide agitation, loss of weight etc. The clinician rates the responses to each question by interviewing the patient and observing the symptoms. Depending on the questions 3–5 possible answers are available to rate the severity of the depression. Interpretation of the total score varies according to different recommendations (e.g. 0–7 = normal; 8–13 = mild depression; 14–18 = moderate depression; 19–22 = severe depression and ≥23 = very severe depression).

The test-scale, the ICD-10 symptom scale for depression (ICD-10-D) was derived 1:1 from the diagnostic criteria for depression given in the ICD-10 chapter of affective disorders (depressive syndrome, F32–F34) [19] to generate a symptom scale. The 10 criteria employed for construction were lowering of mood, loss of interest, lack of energy, reduction of self-esteem and self-confidence, ideas of guilt or worthlessness, suicidal thoughts, difficulty in concentrating or thinking, psychomotor retardation or agitation, disturbed sleep, altered appetite (table 1). For severity rating, a 5-point Likert scale was used ranging from 0 = absent to 4 = very severe for each symptom. For the ICD-10-D the total sum score was used.

Table 1. ICD-10-D and patients' rating (%)

Symptoms	Rating, %				
	0 = absent	1 = mild	2 = moderate	3 = severe	4 = very severe
Lowering of mood	0.9	11.1	57.1	31.0	0.0
Loss of interest	9.3	23.6	42.2	24.0	0.9
Lack of energy	9.3	23.0	46.9	20.8	0.0
Reduction of self-esteem and self-confidence	17.7	15.0	38.9	28.3	0.0
Ideas of guilt or worthlessness	24.8	28.3	29.2	17.7	0.0
Suicidal thoughts	66.8	20.4	6.2	6.2	0.4
Difficulty concentrating or thinking	13.7	38.5	31.4	15.5	0.9
Psychomotor retardation or agitation	33.2	31.9	24.3	10.2	0.4
Disturbed sleep	6.7	23.1	39.6	30.7	0.0
Altered appetite	33.3	28.9	25.8	11.6	0.4

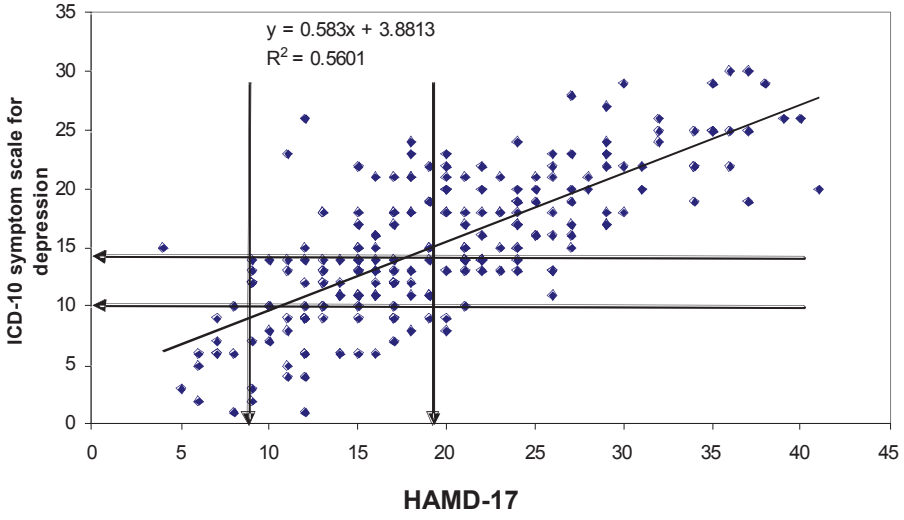

Fig. 1. Correlation between HAMD-17 and ICD-10 sum score at admission (ρ = 0.75, p < 0.0001).

To differentiate the depressive symptoms of the patients, we examined other common psychiatric co-morbidities often related with depression (i.e. somatic symptoms of anxiety/depression, somatoform disorder, and anxiety).

Statistics

Statistical analysis was performed with WinSTAT Version 2001.1 for Windows (SPSS-validated). The continuous data are presented as means, standard deviations (SD); additionally medians as well as 95% confidence intervals and numbers of patients were calculated. Categorical data are presented using counts and percentages rounded to one decimal place. All p-values are 2-tailed, and $p < 0.05$ was considered statistically significant.

Validity of the ICD-10 scale was examined by calculating the correlation of the total score with the HAMD-17. For the correlation coefficient ρ ranging from –1 to 1, correlations between 0.5–1.0 were interpreted as strong. For the co-morbidities ρ was also calculated.

Sensitivity and specificity of a quantitative test depend on the cut-off value above or below which the test is positive. In general, the higher the sensitivity, the lower the specificity, and vice versa.

Sensitivity = True positives / True positives + False negatives (1).
Specificity = True negatives / True negatives + False positives (2).

The testing of sensitivity and specificity of increasing lower and upper cut-offs of the ICD-10 scores versus the HAMD-17 as reference were calculated manually, and the best trade-offs between specificity and sensitivity were chosen as cut-offs. The best trade-off between specificity and sensitivity was identified as the one yielding the highest value when calculating the product of specificity and sensitivity. Graphically, plotting specificity and sensitivity by increasing cut-offs, this point corresponds to the intersection of the 2 curves. For the current analysis we have employed the recommended cut-offs for the HAMD-17: lower limit = 8 and upper limit = 18 [12, 13].

Results

Of the 226 patients recruited (173 females, 46 males, 7 no data), 219 provided full sets of data and could be analyzed. In median their age was 49 years (mean ± SD 50.52 ± 14.70 years) and they had a median duration of symptoms of 3 months (13.88 ± 28.07 months). Their main demographic data are summarized in table 2.

Table 2. Demographics of participating patients

		N	%
Working status	homework	62	27.4
	part-time	31	13.7
	full-time	61	27.0
	retired	42	18.6
	unemployed	24	10.6
	no data	6	2.7
	total	226	100.0
Duration of symptoms, months	< 1	57	25.2
	1 to 3	60	26.5
	3 to 12	55	24.3
	12 to 24	25	11.1
	24 to 240	28	12.4
	no data	1	0.4
	total	226	100.0
Age, years	< 20	4	1.8
	20 to 35	35	15.5
	35 to 50	74	32.7
	50 to 65	65	28.8
	65 to 80	46	20.4
	no data	2	0.9
	total	226	100.0

For validation of the ICD-10 scale n = 219 patients could be analyzed. The ICD-10 scale for depression showed a significant and strong correlation with the HAMD-17 ($p < 0.0001$; ρ = 0.75; fig. 1). Yet, ρ was weaker in the sub-scores HAMD 1–7 including the depression triad (depressed mood, guilt, suicidal thought) [10] and 8–17 including symptoms with an often co-morbid impact (e.g. anxiety, somatic, hypochondriac symp-

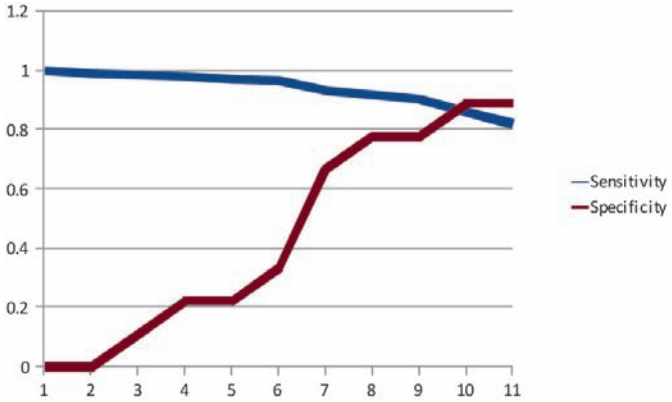

Fig. 2. Testing sensitivity, specificity by increasing lower cut-offs of ICD-10 scores. HAMD = reference, performance of ICD-10 criteria chosen.

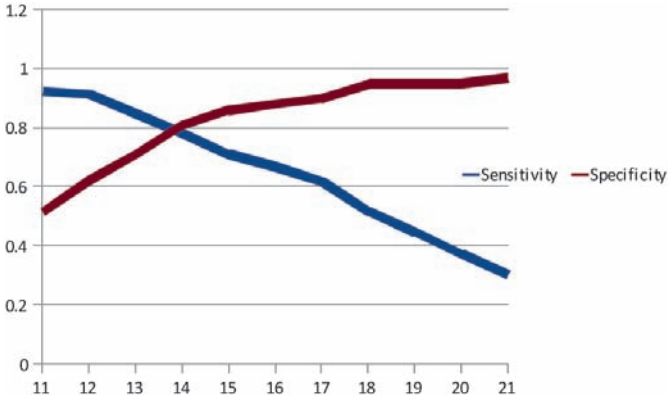

Fig. 3. Testing sensitivity, specificity by increasing upper cut-offs of ICD-10 scores. *HAMD* = reference, performance of ICD-10 criteria chosen.

toms; table 3). However, in terms of psychiatric co-morbidities analyzed ρ showed a rather good discrimination in favor of depression (depression ρ 0.75 > somatic symptoms of anxiety/depression ρ = 0.66 > somatoform disorder ρ = 0.53 > anxiety ρ = 0.20).

The findings of the Clinical Global Impression Scale (CGI) rated by physicians as well as patients were in line with the findings of the ICD-10 and HAMD and have already been published just as the data for the Visual Analogue Scale (VAS) which correlated much weaker [18].

Testing for sensitivity and specificity by increasing lower and upper cut-offs of the tested ICD-10 scores versus HAMD-17 scores is shown in figure 2 and 3. The best trade-offs between specificity and sensitivity of the ICD-10 score were found at 10 points for the lower and 14 points for the upper cut-off (mild depression).

65.3% of the patients were correctly classified by the ICD-10 scale (table 4). However, the severity of the disorder was overestimated in 11.4% and underestimated in 23.3% of the patients. The overall sensitivity and specificity was 76.7% and 88.6%.

Table 3. Performance of ICD-10 criteria for depression chosen in terms of correlation with HAMD-17

HAMD (n = 219)	ICD-10
1–17	
ρ	0.75
p-value	< 0.0001
1–8	
ρ	0.74
p-value	< 0.0001
9–17	
ρ	0.65
p-value	< 0.0001

Table 4. Agreement between individual ICD-10 and HAMD-17 scores (N and % of patients)

ICD-10 / HAMD-17	< 8	Within 8 to 18	> 18
< 10	8 (3.7%)[a]	29 (13.2%)[b]	3 (1.4%)[b]
Within 10 to 14	0 (0.0%)[b]	49 (22.4%)[a]	19 (8.7%)[b]
> 14	1 (0.5%)[b]	24 (10.9%)[b]	86 (39.2%)[a]

[a]Scores agree.
[b]Scores disagree.

Table 5. Lower and upper cut-offs

	Lower cut-offs	
ICD-10/HAMD-17	>8	<8
>10	82.41%[a]	0.46%[b]
<10	13.43%[b]	3.70%[a]
	Upper cut-offs	
ICD-10/HAMD-17	>18	<18
>14	39.81%[a]	11.57%[b]
<14	10.19%[b]	39.81%[a]

[a]Scores agree.
[b]Scores disagree.

Regarding the lower limit (i.e. distinguishing between depression and normal) sensitivity was at 0.86 and specificity at 0.89. However, the relevance of this assertion is limited by the small number of patients in the normal range in both the HAMD-17 and the ICD-10 scale (table 5). For the upper limit (i.e. distinguishing between mild to moderate depression and severe/very severe depression) sensitivity was at 0.80 and specificity at 0.77.

Discussion

In daily medical practice of the medical systems in the Northern hemisphere, physicians are confronted with different aspects of EbM in order to contribute to best possible clinical practice which primarily ought to serve the patient and secondly might limit health costs. EbM provides a basis to choose the right procedures which are necessary for a valid diagnosis of illnesses and efficacious treatment of patients. As health

costs seem to dominate not only political decisions in the medical system but also the choice of medical action (e.g. treatment, diagnosis), easily available diagnostic instruments become a precious value in the course of shared decision-making and treatment [14].

Consequently, scales of diagnostic criteria that are already known to physicians, such as the ICD-10 criteria for depression which are generally used for documentation and referral, seem appropriate when they are valid in comparison with widely accepted scales such as the validated HAMD-17. To match both is highly rational and of great practical relevance. However, there is an ongoing discussion which depression scale could be seen as a gold standard [12, 13]. Additionally, the debate on inter-rater reliability accounts for the different valid questionnaires [6] and partly seems to be connected with clinical experience of physicians as such [8, 20]. This is a proof of the empirical knowledge on which medicine is based next to research evidence. Therefore, the idea to use a rating scale like the generated ICD-10-D seems reasonable in terms of the fact that physicians working in primary care are familiar with the ICD-10 itself. Moreover, the suitability of the ICD-10 criteria for depression have been examined before among GPs [16] and have been found to be sufficiently stable. Only the diagnostic stability over time is discussed although this seems arbitrary and related to other factors than validity itself [21]. ICD-10 criteria for depression also have been used in validation studies for other scales, like the self-rated Hopkins Symptom Checklist-10 (HSCL-10) [9] or the self-rated ICD-10 symptom rating (ISR) [7, 22].

Nevertheless, there are limitations of this study. First of all, the study took place in a mixed physician group. One might argue that it would have been crucial to analyze inter-rater differences. On the one hand, we could not analyze this in the current preliminary study because patients were examined by one rater only. On the other hand, the ICD-10 criteria for depression have already shown to have a good inter-rater reliability in differentiating between depressed and non-depressed patients [23]. Secondly, the validation population is rather small and of limited variance of the patients' condition as can be seen in table 4. Noteworthy, it has been pointed out that making a diagnosis for a condition would depend both on the discriminatory value of the test and the prevalence of the disease in the population of interest [24]. Under this premise the preliminary validation data of the examined ICD-10 scale seem fair concerning correlation, sensitivity and specificity in the population tested, and for the feasibility purpose of the study. But one would have to admit that the findings need to be validated in a rigorous validation study. Thirdly, one might point out that our population consisted of inpatients only. But the difference between in- and outpatients might be more of scientific interest as they should not differ in depressive symptoms as such but in severity, which one would expect to be higher among inpatients. As a matter of fact, the current results can hardly be extrapolated on severely depressed patients.

The discussion about the use of the HAMD-17 [12, 13] shows a need not only for valid but also for widely accepted diagnostic instruments. The potential benefit of an ICD-10-derived symptom scale could be that it provides non-psychiatrists/GPs with an instrument that uses rather non-specialized termini which they are already trained to use as diagnostic criteria within the ICD-10 and which they ought to use for documentation and inter-physician communication anyway. Interestingly, a different ICD-10-derived scale, the ISR, has been developed and validated as a diagnostic tool for a variety of psychiatric disorders [7, 22]. For depression it is quite similar in terms of the 3 main symptoms of depression also covered by our scale. Thus, our scale (10 items) takes a kind of middle position in the number of depressive symptoms rated in the HAMD-17 (17 items) and the ISR (3 items). This underlines the need for evaluating an additional test instrument based on a generally accepted range of symptoms as well as being widely known and freely distributed and available. This should be a future goal not only for the rating of mental disorders [25] but also for daily practice in our Western medical system itself in terms of rational medical care and easily available evaluation instruments for diagnosis and treatment response.

Conclusion

The ICD-10 symptom scale examined in the current population was found to have fair correlation with the HAMD-17 as well as, in face of the limited variance of the patients' condition, acceptable sensitivity and specificity. However, a rigorous validation trial is necessary to corroborate these findings.

Acknowledgements

The observational study was supported by an unconditional grant from Bionorica, Germany whereas this study did not receive any grant. We thank the investigators and patients who participated in the study.

Disclosure Statement

The authors declare no conflict of interest.

References

1 Rush AJ: Limitations in efficacy of antidepressant monotherapy. J Clin Psychiatry 2007;68:8–10.
2 Angst J: How recurrent and predicatable is depressive illness; in Montgomery S, Rouillon F (eds): Longterm Treatment of Depression. Chichester, Wiley, 1992, pp 1–13.
3 Pezawas L, Angst J, Kasper S: Recurrent brief depression revisited. Int Rev Psychiatry 2005;17:63–70.
4 Mitchell A, Vaze A, Rao S: Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 2009;374:609–619.
5 Gilbody S, Sheldon T, House A: Screening and case-finding instruments for depression: a meta-analysis. Can Med Assoc J 2008;178:997–1003.
6 Williams JBW, Kobak KA, Bech P, Engelhardt N, Evans K, Lipsitz J, Jason OB, Pearson J, Kalali A: The GRID-HAMD: standardization of the hamilton depression rating scale. Int Clin Psychopharmacol 2008;23:120–129.
7 Tritt K, von Heymann F, Zaudig M, Zacharias I, Sollner W, Loew T: Development of the ICD-10-symptom-rating (ISR) questionnaire (in German). Z Psychosom Med Psyc 2008;54:409–418.
8 Schmidtke A, Fleckenstein P, Moises W, Beckmann H: Studies of the reliability and validity of the German version of the Montgomery-Asberg Depression Rating Scale (MADRS) (in German). Schweiz Arch Neurol Psychiatr 1988;139:51–65.
9 Haavet OR, Sirpal MK, Haugen W, Christensen KS: Diagnosis of depressed young people in primary health care – a validation of HSCL-10. Fam Pract 2011;28:233–237.
10 Hamilton M: A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62.
11 Beck A, Ward C, Mendelson M, Mock J, Erbaugh J: An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–571.
12 Knappe S, Runge J, Beesdo K, Jacobi F, Wittchen HU: Diagnosing mental disorders: gold or tin standard? – critical comments on standardized diagnostic interviews and clinical routine diagnoses (in German). Psychother Psychosom Med Psychol 2008; 58:72–75.
13 Kriston L, von Wolff A: Not as golden as standards should be: interpretation of the Hamilton Rating Scale for Depression. J Affect Disord 2011;128: 175–177.
14 Hartzband P, Groopman J: Money and the changing culture of medicine. N Engl J Med 2009;360: 101–103.
15 Christensen K: Case-finding and risk-group screening for depression in primary care. Scand J Prim Health Care 2011;29:80–84.
16 Pedersen SH, Stage KB, Bertelsen A, Grinsted P, Kragh-Sorensen P, Sorensen T: ICD-10 criteria for depression in general practice. J Affect Disord 2001;65:191–194.
17 Hiller W, Dichtl G, Hecht H, Hundt W, Vonzerssen D: Testing the comparability of psychiatric diagnoses in ICD-10 and DSM-III-R. Psychopathology 1994;27:19–28.
18 Melzer J, Brignoli R, Keck ME, Saller R: A hypericum extract in the treatment of depressive symptoms in outpatients: an open study. Forsch Komplementmed 2010;17:7–14.
19 Dilling H, Freyberger HJ (eds): Taschenführer zur ICD-10-Klassifikation psychischer Störungen, ed 4. Bern, Huber, 2008.
20 Muller MJ, Dragicevic A: Standardized rater training for the Hamilton Depression Rating Scale (HAMD-17) in psychiatric novices. J Affect Disord 2003;77:65–69.
21 Kessing LV: Diagnostic stability in depressive disorder as according to ICD-10 in clinical practice. Psychopathology 2005;38:32–37.
22 Fischer HF, Tritt K, Klapp BF, Fliege H: Factor structure and psychometric properties of the ICD-10-symptom-rating (ISR) in samples of psychosomatic patients. Psychother Psychosom Med Psychol 2010;60:307–315.
23 Hiller W, Dichtl G, Hecht H, Hundt W, Mombour W, Vonzerssen D: Evaluating the new ICD-10 categories of depressive episode and recurrent depressive disorder. J Affect Disord 1994;31:49–60.
24 Lalkhen A, McCluskey A: Clinical tests: sensitivity and specificity. Educ Anaesth Crit Care Pain 2008; 6:221–223.
25 Mas-Exposito L, Amador-Campos JA, Gomez-Benito J, Lalucat-Jo L: The World Health Organization Quality of Life Scale Brief Version: a validation study in patients with schizophrenia. Qual Life Res 2011;20:1079–1089.